1-1-2014

PKa cycling of the general acid/base in glycoside hydrolase families 33 and 34

Haibo Yu
University of Wollongong, hyu@uow.edu.au

Thomas M. Griffiths
University of Wollongong, tmg994@uowmail.edu.au

PKa cycling of the general acid/base in glycoside hydrolase families 33 and 34

Abstract

Glycoside hydrolase families 33 and 34 catalyse the hydrolysis of terminal sialic acid residues from sialyl oligosaccharides and glycoconjugates with a net retention of the stereochemistry at the anomeric centre. It is generally believed that the conserved aspartic acid in the active site functions as a general acid to protonate the hydroxyl group of the departing aglycone during glycosylation, and then as a general base to facilitate the nucleophilic attack of the water molecule on the intermediate state during the deglycosylation reaction. The dual role of the general acid/base places specific demands upon its protonation state, and thus pKa values. However, it is not fully understood how this catalytic residue can achieve such pKa cycling during catalysis. We present both MM and combined QM/MM simulations to characterise the pKa values of the proposed catalytic general acid/base in the glycoside hydrolase families 33 and 34. Collectively, our study suggests that the binding of anionic substrates and the local solvation properties along with the neutralisation of the nearby glutamic acid upon glycosylation modulate the electrostatic environment around the general acid/base to achieve its proper protonation states.

Keywords

CMMB

Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Yu, H. & Griffiths, T. M. (2014). PKa cycling of the general acid/base in glycoside hydrolase families 33 and 34. Physical Chemistry Chemical Physics, 16 (12), 5785-5792.

ARTICLE TYPE | XXXXXXXX

pK_a cycling of the general acid/base in glycoside hydrolase families 33 and 34†

Haibo Yu*[a] and Thomas M. Griffiths[a]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Glycoside hydrolase families 33 and 34 catalyse the hydrolysis of terminal sialic acid residues from sialyl oligosaccharides and glycoconjugates with a net retention of the stereochemistry at the anomeric centre. It is generally believed that the conserved aspartic acid in the active site functions as a general acid to protonate the hydroxyl group of the departing aglycone during glycosylation, and then as a general base to facilitate the nucleophilic attack of the water molecule on the intermediate state during the deglycosylation reaction. The dual role of the general acid/base places specific demands upon its protonation state, and thus pK_a values. However, it is not fully understood how this catalytic residue can achieve such pK_a cycling during catalysis. We present both MM and combined QM/MM simulations to characterise the pK_a values of the proposed catalytic general acid/base in the glycoside hydrolase families 33 and 34. Collectively, our study suggests that the binding of anionic substrates and the local solvation properties along with the neutralisation of the nearby glutamic acid upon glycosylation modulate the electrostatic environment around the general acid/base to achieve its proper protonation states.

1 Introduction

Glycoside hydrolases (GHs, or glycosidases) catalyse the hydrolysis of carbohydrates and play a variety of key roles in biological processes. They have many practical applications in medicine and biotechnology [1]. GH can be categorised into two distinct mechanistic classes: retaining and inverting GHs, depending on the stereochemistry of the hydrolysis product with respect to that of the reactant [2–6]. The insightful 1953 paper by Koshland hypothesized the molecular mechanisms for both retaining and inverting GHs [7]. Inverting GHs operate via a direct displacement of the leaving group by water employing a pair of carboxylic acids at the active site with one serving as the general acid and the other one serving as the general base. In contrast, retaining GHs catalyse hydrolysis via a double-displacement mechanism. In the first step, a general acid catalyst (almost always Asp or Glu) facilitates the departure of the leaving group. A second residue acts as nucleophile (most commonly a carboxylate residue) to stabilise an enzyme sequestered intermediate state. The second step of the reaction would be the near microscopic reverse of the first step; the general base facilitates the attack by a lytic water on the anomeric centre to break down the intermediate state, forming the product with retained stereochemistry at the anomeric centre. With typical rate enhancements of up to 10^{17}-fold, GHs are among the most proficient enzymes; thus, understanding the structural, dynamic, and electrostatic bases underlying the mechanisms remains the subject of considerable scientific effort.

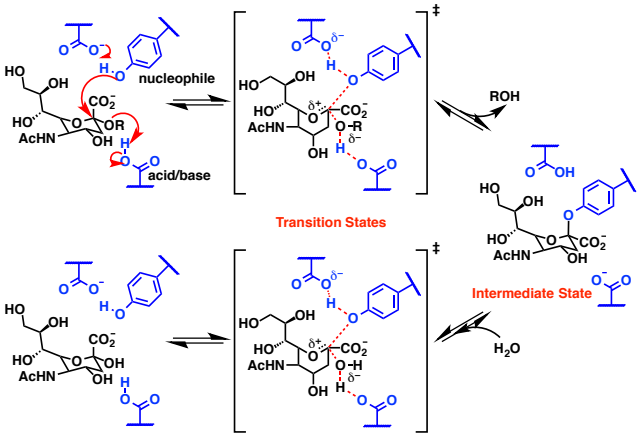

Fig. 1 The proposed catalytic mechanism for GH33 and 34.

Sialic acid is the common name for *O*- and *N*-substituted derivatives of a nine carbon monosaccharide called neuraminic acid [8]. These acids lie on the terminal position of cell surface glycoproteins and glycolipids and play an important role in molecular recognition by the immune system [9]. GH33 and 34 catalyse the hydrolysis of sialosides with a net retention of the stereochemistry at the anomeric centre. It has

[a] *School of Chemistry, University of Wollongong, Wollongong NSW 2552, Australia. Fax: +61 2 4221 4287; Tel: +61 2 4221 4235; E-mail: hyu@uow.edu.au*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

been shown that both families use a tyrosine/glutamate couple as the catalytic nucleophile, rather than the more common single glutamate found in the majority of the retaining GHs (Figure 1). It has been rationalised that this is most likely due to the chemical nature of the substrate, sialic acid, which possesses a carboxylate group adjacent to the anomeric centre. This would introduce unfavourable electrostatic interactions between the negatively charged nucleophile and the carboxylate group. By employing a tyrosine/glutamate nucleophilic pair, the tyrosine residue, deprotonated by the nearby glutamate residue, is able to perform nucleophilic attack without facing such repulsive electrostatic interactions (Figure 1). This mechanism is supported by the recent captured covalent intermediate state for both *Trypanosoma cruzi* trans-sialidase (TcTS, GH33) and influenza neuraminidase (NA subtype N9, GH34) using either an enzyme mutant or a substrate that exhibits slow turnover [10–12]. *Trypanosoma cruzi* trans-sialidase is a key enzyme associated with infection by *Trypanosoma cruzi*, which causes the tropical parasitic disease, Chagas disease. TcTS catalyses the transfer of sialic acids from mammalian host cells to parasitic cell surfaces in order to escape detection by the immune system. It has been proposed as a promising target for the development of therapeutics to treat Chagas disease [13]. Influenza neuraminidase is a proven anti-influenza drug target. Zanamivir and oseltamivir, among the four licensed prescription influenza antiviral agents available, target NA [14]. Considering their great medical relevance, both enzymes have been subjected to extensive experimental studies [15, 16]. Complementary combined quantum mechanics/molecular mechanics (QM/MM) simulations demonstrated that the substrate binding significantly shifted the pK_a value of the nucleophile Tyr342 in TcTS - enabling the proton transfer from the hydroxyl group of a tyrosine to a carboxylate group [17]. QM/MM free energy studies of the catalytic mechanism of TcTS confirm the experimental findings that the reaction processes through a long-lived covalent intermediate state [18].

It is generally believed that the conserved Asp59 in TcTS or Asp151 in NA functions as both a general acid and a general base during the catalysis in the double-displacement mechanism (Figure 1). Such a dual catalytic role, accompanying the so-called "pK_a cycling", requires that the enzymes are able to control its protonation state at each step of the catalysis: from the apo-state, to the Michaelis complex state, to the intermediate state. pK_a cyclings are well-precedented for acid/base-catalysed hydrolysis of glycoside [19–21]. Numerous studies have focused on dissecting the molecular factors that establish the pK_a values of the catalytic residues in retaining GHs including hen gee white lysozyme [22, 23] and *Bacillus circulans* xylanase [19, 21]. It has been postulated that pK_a cycling is likely to be a common phenomenon in retaining GHs. A full understanding of the structural and energetic basis for pK_a cycling is thus essential for a complete description of their catalytic mechanisms. It has been noted that in the apo structures Asp59 in TcTS (PDB code: 1MS3 [24]) or Asp151 in NA (PDB code: 2QWA [25]) might not function as an acid catalyst since it is largely solvent exposed and hence it may have a pK_a too low to allow it to serve such a role under physiologically-relevant pHs. On the other hand, based on the the Michaelis complex obtained for TcTS (PDB code: 1S0I [26]), one might argue that the position of the aromatic side chain of Tyr119 and the binding of the aglycone of the substrate will significantly decrease the solvent exposure of Asp59. Furthermore, the anionic nature of the substrate itself (the carboxyl group in sialosides) might raise the pK_a of Asp59 even further.

A key question prompted is what molecular features in GH33 and 34 establish the pK_a values of the general acid/base in the free, Michaelis complex and glycosyl-enzyme form. Complementary to experimental studies, combined QM/MM simulations has been proven to be a powerful alternative in revealing mechanistic insights. In computational enzymology, a major challenge is to develop computational algorithms to accurately describe the highly heterogeneous electrostatic environment within these macromolecules, and to delineate the factors modulating the precise pK_a values of ionisable groups in enzymes, as required to identify the catalytically competent protonation state and understand the catalytic mechanisms [27]. Their protonation states are determined by the intrinsic pK_a of the functional groups and the microenvironment created by the enzyme structure, which often perturbs their intrinsic pK_as to their so-called functional or apparent pK_as. Such pK_a shifts can be achieved by interactions between the catalytic group with other charged or polar groups, or by the polarity or dielectric of the medium that surrounds it [28]. From a computational point of view, it is very challenging for methods purely relying on continuum electrostatics or empirical models to capture the variation in pK_as involving substantial structural and energetic changes in a heterogeneous system. Microscopic simulations based on MM or QM/MM with explicit solvent representations, therefore, are preferred in these challenging cases where the explicit solvation and structural reorganisation need to be properly taken into account. Such simulations bring additional insights into the mechanistic modulation of the pK_as for catalytically important residues.

In the current work, we apply different computational methods to calculate the pK_a shifts of the general acid/base Asp59 in TcTS of GH33 and Asp151 in NA of GH34. There are two main goals for this study. First, we hope to reconcile the proposed roles for Asp59 in TcTS and Asp151 in NA with their protonation states at different catalytic steps. Second, we would like to explore the microscopic origins for such pK_a shifts in these two enzymes, and more broadly, in the families of GH33 and 34.

2 Computational details

2.1 Structure preparation and classical MD simulations

All the simulations were carried out using the program CHARMM (version c38a1) [29]. The protein and substrate atoms are described with the all-atom CHARMM force field c36 [30] and the water molecules are described with the TIP3P model [31]. The efficient General Solvent Boundary Potential (GSBP) setup [32] was adopted to simulate a spherical region of 20 Å radius centred on the carboxyl group of interest with explicit water molecules. Water molecules are added to the system following the standard protocol of superimposing the system with a water sphere. In all cases, the system is partitioned into a 20 Å inner region, while the rest of the protein is treated as the outer region within the GSBP framework. The long-range interactions were taken into account in the form of the solvent-shielded static field and the solvent-induced reaction field via 196 harmonic functions in the multipole expansion. The solvent-shielded static field and the reaction-field matrix were calculated once with finite-difference Poisson-Boltzmann (PB) equation, assuming a dielectric constant of 1.0 inside the protein, immersed in a solvent with a dielectric constant of 78.5. The atomic Born radii for protein atoms used to set up the dielectric boundaries in the PB calculations were optimised by free energy perturbation/molecular dynamics (FEP/MD) simulations with explicit solvent models [33]. Non-bonded interactions within the inner sphere are treated with an extended electrostatics model, in which charge groups beyond 12 Å interact as multipoles. The water molecules within the inner region were confined by a non-polar cavity potential to prevent entry into the surrounding dielectric continuum. Newton's equations of motion are solved for the MD region (within 18 Å), and Langevin equations of motion are solved for the buffer region (18-20 Å) with a temperature bath of 298.15 K. All the bonds involving hydrogen atoms are kept rigid with the SHAKE algorithm [34] and the time-step is set to 1.0 fs.

As starting structures for the simulations, we used the X-ray structures of TcTS (PDB code: 1MS3 [24], resolution 1.65 Å and 1S0I [26], resolution 1.60 Å) and NA (subtype N9, PDB code: 2QWB [25], resolution 2.00 Å and 3W09 [12], resolution 2.00 Å). In the case of the structures solved for an inactive mutant, the residue was mutated back to the wild-type sequence. All the simulated systems are summarised in ESI Table 1.

2.2 Combined QM/MM MD simulations

An efficient and approximate density function theory (DFTB3) [35] was adopted as the QM method in the combined QM/MM simulations [36–41]. DFTB3 is a recent extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB) [42, 43] and derived from a third order expansion of the density functional theory total energy around a given reference density [35]. It has been shown to substantially improve the description of charged systems, especially regarding hydrogen binding energies and proton affinities. As a result, DFTB3 is particularly applicable to biomolecular systems. As shown by benchmark calculations on proton affinities using G3B3 as reference (for details, see ESI Table 4), DFTB3 is appropriate for the current system of interest. The QM/MM boundary (between the Cα and Cβ in Asp or the Cβ and Cγ in Glu and Tyr) is treated using link-atoms with the divided frontier charge scheme (DIV as implemented in CHARMM) [44]. The starting structures for the QM/MM simulations are taken from the equilibrated MM simulations. The setup details are identical to the MM simulations. Unless explicitly stated, the QM region includes the side chain of Asp59 in TcTS or Asp151 in NA (ESI Table 1).

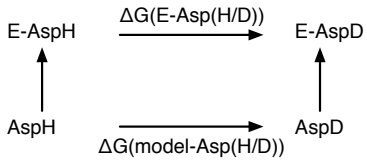

Fig. 2 The thermodynamics cycle used to calculate the pK_a shifts for the general acid/base in GH33 and 34 with explicit solvent simulations. The upper leg represents the free energy difference when perturbing the proton (H) into a dummy atom (D) for the aspartic acid in a protein (E). The lower leg represents the free energy difference when perturbing the proton (H) into a dummy atom (D) for an isolated aspartic acid with neutral blocking groups in aqueous solution. The double free energy difference between the two legs is defined as
$\Delta\Delta G = \Delta G(\text{E-Asp(H/D)}) - \Delta G(\text{model-Asp(H/D)})$.

2.3 pK_a calculations

The thermodynamics cycle used to calculate pK_a shifts, *i.e.* the pK_a difference between an aspartic acid side chain in a protein and the same group in solution (aspartate), is shown in Figure 2 [45, 46]. The pK_a shifts are then calculated according to Equation 1.

$$\begin{aligned} \Delta\text{p}K_a &= \text{p}K_a(\text{E-Asp}) - \text{p}K_a(\text{model-Asp}) \\ &= \frac{1}{2.303kT}\Delta\Delta G \end{aligned} \quad (1)$$

The apparent pK_a (pK_a(E-Asp)) can be calculated based on the experimental data for pK_a(model-Asp) of 3.8.

The pK_a calculations were first carried out with either the continuum electrostatics based method (PB) or an empirical function based method (PROPKA) [47]. The PB calculations

were carried out for the X-ray structures with the PBEQ module within CHARMM [29]. The calculations were performed in two steps, first using a coarse grid spacing of 1.50 Å followed by a focusing around the main region with a finer grid spacing of 0.25 Å centred at the ionisable group of interest. The atomic Born radii optimised by explicit solvent FEP/MD simulations were adopted [33]. The free energy differences (ΔG(E-Asp(H/D)) and ΔG(model-Asp(H/D))) were obtained by solving the PB equations both in the gas phase (with the dielectric constant set to 1.0) and in the aqueous solution (with the dielectric constant of water set to 78.5). The dielectric constant of protein was to set to either 4.0 or 8.0 and the salt concentration was set to 0.150 M. In contrast, PROPKA adopts a very fast empirical energy function that takes into account the desolvation effects and intra-protein interactions and the most recent version has also been shown to provide a reliable treatment for the coupling between the protein and its ligands, which is crucial for the current study[47].

For explicit solvent representation, the pK_a calculations were carried with both MM and combined DFTB3/MM simulations using a thermodynamic integration approach within the Dual Topology Single Coordinate scheme [45]. Such simulations have been used in several different systems and have been shown to successfully reproduce the experimental pK_a shifts of small molecules in solutions and (semi)-buried ionisable residues in proteins [48–51]. The free energy differences (ΔG(E-Asp(H/D)) and ΔG(model-Asp(H/D))) are determined by integrating the converged free energy derivatives ($\delta G/\delta\lambda$) over the coupling parameter λ from 0 (the protonated state) to 1 (the deprotonated state). Specifically, 11 evenly spaced windows between 0 and 1 were adopted in the current simulations. To carefully monitor the statistical and sampling errors and reproducibility of the microscopic pK_a simulations, multiple independent trajectories have been carried out by starting the FEP simulations from different structures collected from equilibrium MM and QM/MM simulations (See ESI Table 5 and 6 for more details). As a reference, ΔG(model-Asp(H/D)) is estimated to be -47.3 kcal/mol with the MM simulation and 121.2 kcal/mol with the DFTB3/MM simulation, respectively.

3 Results and Discussions

3.1 pK_a calculations based on the Poisson-Boltzmann equation and PROPKA

The pK_a values of the proposed general acid Asp59 in TcTS and Asp151 in NA were predicted based on either the Poisson-Boltzmann equation implemented in the PBEQ module in CHARMM [29] or the empirical method PROPKA 3.1 [47] (Table 1). These calculations consistently indicate that Asp59 in TcTS and Asp151 in NA should be deprotonated in the apo state (1MS3-apo, 1S0I-apo and 2QWB-apo) and the intermediate state (3W09-int), despite numerical differences between different methods (PB vs. PROPKA) and different structures used (1MS3 vs. 1S0I). For the Michaelis complex, the PB calculations predicted that Asp59 in TcTS will be protonated (1S0I-holo) while Asp151 in NA remains deprotonated (2QWB-holo). PROPKA predicted that both remain deprotonated in the Michaelis complex (1S0I-holo and 2QWB-holo), which is not consistent with their proposed role as a general acid in the first step of hydrolysis. Systematic pK_a predictions with PROPKA on all the available structures for GH33 and 34 families failed to capture the proposed pK_a upshifts upon ligand binding (ESI Table 2 and 3). The discrepancy is most likely due to the uncertainties in the pK_a calculations using PBEQ or PROPKA when dealing with residues where protonation or deprotonation is accompanied with substantial change in their local solvation environment (e.g. from largely solvent exposed to deeply buried), as it is seen in other cases [50].

Table 1 The pK_a values of Asp59 in TcTS or Asp151 in NA based on the Poisson-Boltzmann equation (PBEQ) or the empirical method PROPKA.

Structure	PBEQ[a]		PROPKA	Protonation[d]
	$\varepsilon_p = 4$	$\varepsilon_p = 8$		
Asp59 in *Trypanosoma cruzi* trans-sialidase				
1MS3-apo	3.4	3.3	2.2	**Deprot.**
1S0I-apo	4.2	3.9	2.1	**Deprot.**
1S0I-holo[b]	9.6	7.4	2.4	?
1S0I-holo2[c]	8.3	6.7	3.1	?
Asp151 in influenza neuraminidase				
2QWB-apo	4.1	4.1	3.6	**Deprot.**
2QWB-holo[b]	5.1	4.9	3.6	*Deprot.*
3W09-int	4.4	4.3	4.2	**Deprot.**

[a] A reference pK_a value of 3.8 for Asp in aqueous solution was applied. [b] The complexed ligand is lactose sialic acid; [c] The complexed ligand is sialic acid. [d] The predicted protonation states from the PBEQ and PROPKA calculations. If they are consistent with each other, a protonation state is assigned; Additionally if they are consistent with their proposed role in Figure 1, it is shown in bold otherwise in italic. If the two methods are not consistent with each other, it is noted by a ?.

3.2 pK_a calculations based on explicit solvent simulations

To overcome the known limitations of pK_a calculations based on the continuum electrostatics protocol or the empirical method, we carried out additional pK_a shift calculations of Asp59 in TcTS and Asp151 in NA with explicit solvent simulations, using both the classical MM potential and the combined QM/MM potential (Table 2). The final snapshots from these equilibrium DFTB3/MM simulations are shown in Figure 3 and the heavy-atom root-mean-square deviations (RMSD) from the X-ray structure for the active site residues

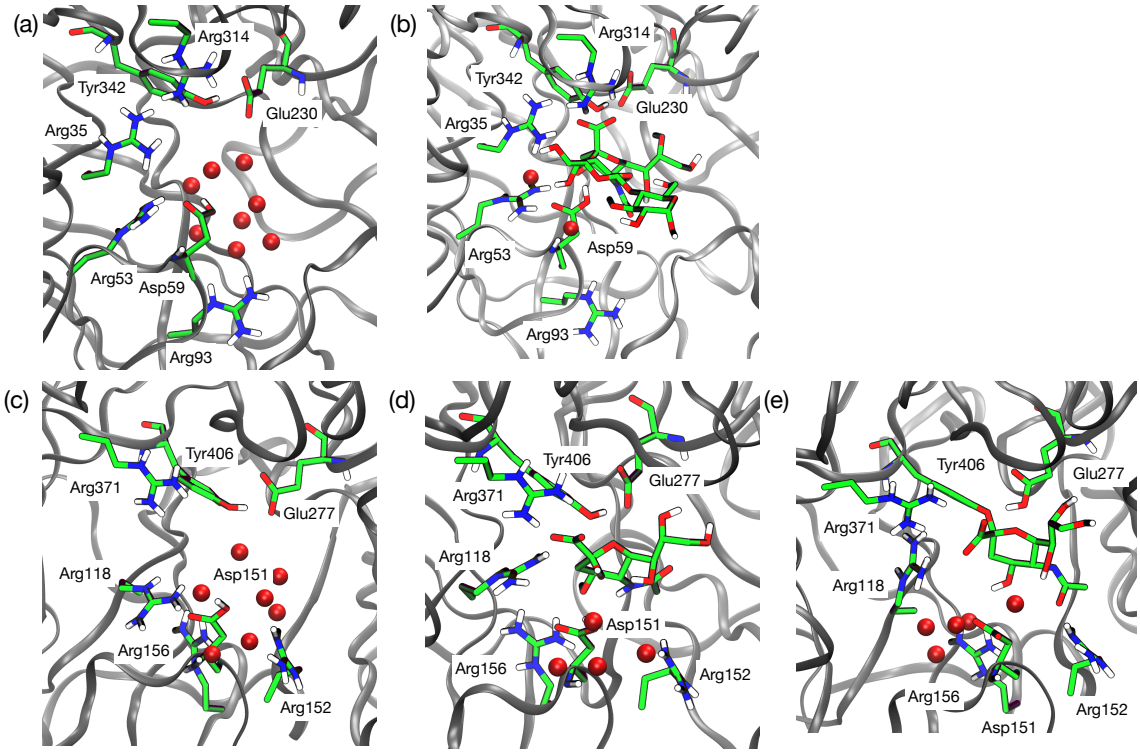

Fig. 3 The final snapshots at 20 ns from the DFTB3/MM simulations. The catalytic Asp59, Glu230 and Tyr342 in TcTS and Asp151, Glu277, Tyr406 in NA are shown explicitly together the substrate and the key nearby residues. The aliphatic hydrogen atoms are omitted for clarity. The water molecules within 4.0 Å of OD2 in Asp59 or Asp151 are shown in red. (a) 1S0I-apo; (b) 1S0I-holo; (c) 2QWB-apo; (d) 2QWB-holo; (e) 3W09-int.

and the substrate are about 0.8–1.2 Å. Three different snapshots from these equilibrium simulations (20 ns, 30 ns and 40 ns in MM and 10 ns, 15 ns and 20 ns in DFTB3/MM) were taken as the initial structures for explicit solvent FEP simulations. Encouragingly, both MM and QM/MM simulations offer a consistent picture of the protonation states of Asp59 in TcTS and Asp151 in NA. In the apo state, Asp59 in TcTS (1MS3-apo and 1S0I-apo) and Asp151 in NA (2QWB-apo) are deprotonated. For Asp59 in TcTS, two different structures were used as the starting conformations in the equilibrium MD simulations (1MS3 and 1S0I) for the apo state. Both setups predicted that Asp59 in TcTS in the apo state has a pK_a well below 7 (on average 3.2 and 1.7 for MM, 0.8 and 1.0 for QM/MM) and this indicates that the results are not very sensitive to the exact structure adopted in the simulations. When the substrate binds, they have a substantial pK_a upshift, and become protonated (1S0I-holo and 2QWB-holo, Table 2), consistent with their proposed role as a general acid in the first step (Figure 1). It is worth noting that there is a semi-quantitative agreement between the MM and QM/MM simulations. The advantage of using a QM/MM rather than a MM potential here is largely due to the convenience of computational setup and our general interest in understanding enzymatic reactions with combined QM/MM methods; the ability to reproduce pK_a shifts serves as a stringent benchmark for QM/MM simulations - especially those involved in proton transfers. For the case of the glycosyl-enzyme intermediate state in NA (3W09-int), only QM/MM simulations were carried out since there is no published CHARMM force field parameters available for the glycosyl-enzyme covalent linkage. The pK_a calculations predict that the pK_a of Asp151 in NA is about 6.3. It is 2.4 pK_a units lower than the 2QWB-holo state. This pK_a shift is smaller than that between the 2QWB-apo state and 2QWB-holo state (4.6 pK_a units with QM/MM), which is expected since the possible structural and energetic perturbations accompanying the formation of the intermediate state is smaller than those induced by the binding of an anionic substrate.

The predicted pK_as from the explicit solvent simulations are consistent with available experimental data. It has been shown that TcTS has an apparent optimal pH of 7.9 [52], under which in the Michaelis complex state (1S0I-holo) Asp59 will be protonated to serve as the general acid (pK_a: 9.6 with MM and 8.5 with QM/MM, Table 2). The predicted optimal

pH range for NA is between 6.3 and 8.7 from QM/MM pK_a calculations, imposed by the required protonation states of the general acid/base Asp151 during the catalysis cycle (Table 2). This is in quite a good agreement with the experimental estimation. NA has a broad pH curve and the optimal pH has been measured to be around 6.0 (e.g. 5.9 for H4N2 [53], 6.4 for H3N2 [54] and 6.5 for H1N1 [55]).

3.3 Perturbative analysis of the general acid/base pK_a

To gain additional insights into the importance of a particular molecular factor modulating the pK_as of the general acid/base in GH33 and 34, we carried out a set of perturbative analyses, in which the energy gap between the protonated state and the deprotonated state is reevaluated for the snapshots from the thermodynamic integration of the QM/MM trajectories, after some of the parameters in the QM/MM coupling model are modified. For instance, the effect from the desolvation effects is evaluated by setting the partial charges on the TIP3P water model to zero, and re-calculating the free energy derivatives (by calculating the new energy gap) in the original trajectories for the different λ windows. Strictly speaking, the effects of individual factors are not additive [56] and the relaxation of the modified environment after such perturbation is neglected. Thus, such a perturbative analysis should be treated only qualitatively. Previous studies suggest that even though such a perturbative analysis tends to overestimate the absolute magnitude of contributions, especially those from charged residues, the relative trends are meaningful. It can provide valuable information about how the environment influences the pK_a of the general acid/base. The results from the perturbative analyses are summarised in Table 3.

As a reference, consider the deprotonation of NH_3-Asp-OCH_3 in solution, the contribution from the surrounding water is about -41.9 kcal/mol. When setting the partial charges on all the water molecules to zero, the net contribution from water in 1S0I-apo and 2QWB-apo are relatively small, about -1.3 kcal/mol and -7.4 kcal/mol. The reorganisation of the water upon deprotonation (which is related with the variation of the water contribution as a function of λ) however, is substantial; in other words water molecules respond significantly to the change in the protonation state of Asp (See ESI Table 7). Interestingly, the water contributions in 1S0I-holo and 2QWB-holo are systematically larger than those in 1S0I-apo and 2WQB-apo. The net contributions are about -47.3 kcal/mol and -41.9 kcal/mol respectively. These aspects of the water contribution are expected, considering that the overall change in charge upon deprotonation is the same between the apo state and the holo state, however, the overall charge of the systems are different for the apo state and the holo state due to the negative charge on the substrate (-1.0 e). The contributions from water can also be rationalised from the solvation structure of Asp59 in TcTS or Asp151 in NA. Water distribution around the carboxylic oxygen in Asp59 or Asp151 from DFTB3/MM are shown in Figure 4. For 1S0I-apo and 2QWB-apo, there are about 6-7 water molecules within the first solvation shell of Asp59 or Asp151 at both λ = 0.0 and λ = 1.0 (with a cutoff of 4.0 Å). This is consistent with the observation that in the apo state Asp59 or Asp151 are largely solvent exposed and the coordination numbers for the protonated and deprotonated isolated Asp in aqueous solution are 7.4 and 8.0, respectively. With the binding of the substrate, the coordination numbers decrease to 1.6 (λ = 0.0) and 3.6 (λ = 1.0) in 1S0I-holo and 4.6 (λ = 0.0) and 6.3 (λ = 1.0) in 2QWB-holo. From this, we can conclude that the substrate binding makes Asp59 or Asp151 less exposed to water and contributes to their observed pK_a upshifts. The same trend is observed in the case of 2QWB-holo when comparing with 2QWB-apo.

In all four cases, the protein atoms make a major contribution, on the order of more than -80 kcal/mol (Table 3). In contrast, the protein reorganisation during deprotonation, which is reflected by the variation of the protein contribution to the free energy as a function of λ, is substantially smaller than water (ESI Table 7). This suggests that the groups making larger contributions to the deprotonation process have limited flexibility, which is reflected in their very similar equilibrated structures (Figure 3). Strikingly, the perturbative contributions from different residues are almost identical between the apo state and the holo state (Figure 5), the exception being residue R118 and R119 in 2QWB which is likely due to the small structural differences between the apo and holo state (Figure 3 c and d). In other words, the difference in the pK_a values between the apo state and the holo state is most likely not due to the protein. As expected from the chemical nature of the substrate, its binding contributes significantly to the pK_a shifts of Asp59 and Asp151, being 48.4 kcal/mol and 40.2 kcal/mol for 1S0I and 2QWB respectively (Table 3). This is largely compensated by the tightly coupled protein and water contributions resulting in a net pK_a upshift of 7.5 and 4.6 for Asp59 and Asp151 with QM/MM simulations, respectively. Taken together the perturbative analysis, we demonstrate that the binding of substrates upshifts the pK_a of Asp59 and Asp151 through both reducing the solvent exposure and the increased electrostatic repulsion.

From the holo state (2QWB-holo) to the covalent intermediate state (3W09-int), a pK_a downshift of 2.4 (from 8.7 to 6.3 with QM/MM) was observed (Table 2). In the glycosyl-enzyme intermediate state, the solvent exposure for Asp151 is similar to those in 2QWB-holo (Figure 4 b, d) and this suggests that the change in the solvation structure might not be the dominant factor. The inclusion of Glu277, Tyr406 and the substrate with DFTB3 in the 3W09-int simulations prevents a comparative perturbative analysis as in Table 3. The contribution from the individual remaining MM protein residues from

Table 2 The pK_a values of Asp59 in TcTS or Asp151 in NA based explicit solvent simulations[a].

Simulations	MM				DFTB3/MM				Protonation[b]
	A	B	C	Ave.[c]	A	B	C	Ave. [c]	
Asp59 in *Trypanosoma cruzi* trans-sialidase									
1MS3-apo	3.4	3.6	2.7	3.2±0.5	1.1	0.4	0.9	0.8±0.4	**Deprot.**
1S0I-apo	1.1	2.2	1.8	1.7±0.6	0.5	1.0	1.4	1.0±0.5	**Deprot.**
1S0I-holo	9.8	10.2	8.9	9.6±0.7	7.8	8.4	9.2	8.5±0.7	**Prot.**
Asp151 in influenza neuraminidase									
2QWB-apo	5.3	5.6	6.0	5.6±0.4	4.2	4.0	4.2	4.1±0.1	**Deprot.**
2QWB-holo	8.4	9.0	9.6	9.0±0.6	8.8	8.3	9.0	8.7±0.4	**Prot.**
3W09-int	-	-	-	-	6.6	6.4	5.9	6.3±0.4	**Deprot.**

[a] A reference pK_a value of 3.8 for Asp in aqueous solution was applied; ΔG(Asp) is estimated to be -47.3 kcal/mol with MM and 121.2 kcal/mol with DFTB3/MM; [b] See Table 1 footnote [d] for more details; [c] The average pK_as from three different sets of simulations. The error bars in pK_a calculations are estimated from the standard deviation of the three sets of calculated values.

the perturbation analysis gives a very similar profile to those from 2QWB-apo and 2QWB-holo, except Glu277 which is treated with QM in the 3W09-int simulations (Figure 5). Another important factor in modulating the pK_a of the general acid/base in retaining GHs is the interaction between the two carboxylic acids in the nucleophile and the general acid/base [27]. In this scenario from the holo state (2QWB-holo) to the covalent intermediate state (3W09-int), it is the neutralisation of Glu277 upon glycosylation, eliminating the negative charge thus changing the electrostatic interactions between Glu277 and Asp151 (Figure 1). The perturbative analysis in the 2WQB-holo simulation shows that protonating Glu277 will decrease the deprotonation free energy for Asp151 by 14.4 kcal/mol (i.e. decrease the pK_a by 10.5 pK_a units) due to the decreased electrostatic repulsion. This is consistent with the experimental data on *Bacillus circulans* xylanase. The pK_a for the corresponding general acid/base Glu172 decreases from 6.7 in the apo state and the Michaelis complex state to 4.2 in the covalent intermediate state (i.e. 2.5 pK_a units) [19]. Mutagenesis study also showed that mutating Glu78 to a neutral glutamine causes the pK_a of Glu172 in the apo state to decrease from 6.7 to 4.2 (2.5 pK_a units) [19]. All this evidence indicates that charge neutralisation, either by glycosylation or mutagenesis, results in a decrease in the pK_a of the general acid/base, facilitating a thermodynamically favourable deprotonation, thus making Asp151 serve as a general base in the second step.

Table 3 Perturbation analysis of the deprotonation free energy of Asp59 or Asp151 with DFTB3/MM simulations[a].

System	Ref.[b]	ΔAll[c]	ΔWat[e]	ΔProt[e]	ΔLig[f]
NH_3 -Asp-OCH_3 in water					
Asp	121.4	-	163.3	-	-
Asp59 in *Trypanosoma cruzi* trans-sialidase					
1S0I-apo	117.2	202.2	118.5	199.4	-
1S0I-holo	127.3	203.3	174.7	220.1	78.9
Asp151 in influenza neuraminidase					
2QWB-apo	121.5	201.4	114.1	211.5	-
2QWB-holo	127.0	199.6	168.9	213.5	86.8

[a] The deprotonation free energy of Asp59 in TcTS or Asp151 in NA were calculated by integrating the free energy derivatives recalculated from the DFTB3/MM FEP trajectories under different QM/MM coupling perturbations. The values are in kcal/mol; [b] Ref. refers to the perturbative analysis on the originally simulated systems while keeping all the parameters untouched; [c] All the partial charges on the protein (inter and outer atoms), water and the substrate were set to zero; [e] All the partial charges on water were set to zero; [e] All the partial charges on the protein (inter and outer atoms) were set to zero; [f] All the partial charges on the substrate were set to zero.

4 Conclusions

For glycoside hydrolase families 33 and 34, based on the high-resolution crystal structure, an Asp in the active site within hydrogen bonding distance of the glycosidic oxygen has been proposed as the acid catalyst for the initial glycosyl-enzyme formation and the same aspartic acid acts as the general base catalyst for nucleophilic attack in the following deglycosylation step. It is intriguing how this aspartic acid is capable of such pK_a cycling, particularly serving as a general acid within a high solvent exposed environment where its pK_a is too low. The current study, utilising the microscopic free energy perturbation with both MM and QM/MM models, demonstrates that in GH33 and 34 at least two molecular factors are at play: the binding of the substrate significantly decreases its solvent exposure, and the electrostatic nature of the anionic substrate itself further substantially raises the pK_a as a consequence of the repulsive electrostatic interactions. The striking difference between GH33 and 34 and other retaining GHs, including lysozyme [22, 23] and xylanase [19, 21], is the characteristic anionic nature of the substrate (sialosides). Upon binding, the negative charge on the substrate, within the close proximity to the general acid/base, provides much stronger per-

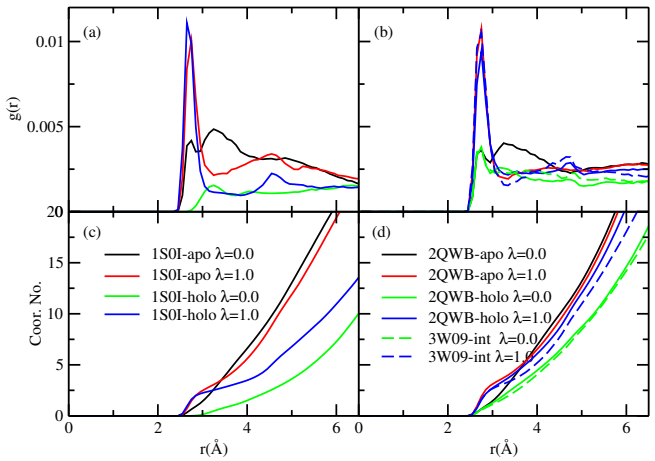

Fig. 4 Water distribution around the carboxylic oxygen in Asp59 in TcTS or Asp151 in NA from the DFTB3/MM simulations. (a) Radial distribution function of water oxygens around the carboxylic oxygen atom in Asp59 in the simulations of 1S0I; (b) Radial distribution function of water oxygens around the carboxylic oxygen atom in Asp151 in the simulations of 2QWB and 3W09; (c) The coordination number for water around the carboxylic oxygen atom in Asp59 in the simulations of 1S0I; (d) The coordination number for water around the carboxylic oxygen atom in Asp151 in the simulations of 2QWB and 3W09.

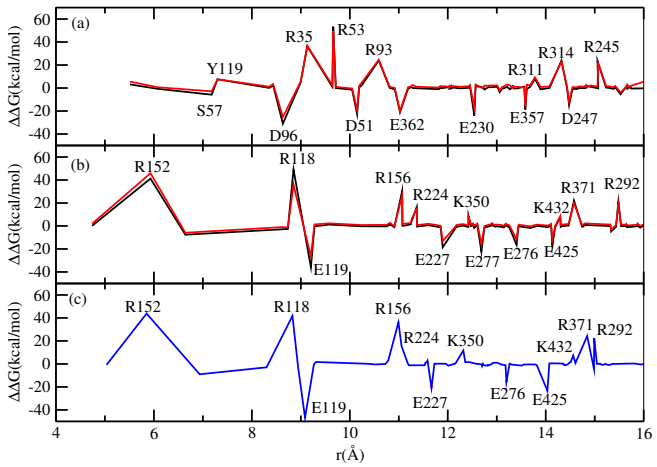

Fig. 5 Contribution to the electrostatic component of the free energy of deprotonation from protein atoms (only MM atoms are considered) in the TcTS and NA from the DFTB3/MM simulations. The contribution is plotted against the distance between its Cα from the Cγ of Asp59 in TcTS or Asp151 in NA. (a): 1S0I-apo (black) and 1S0I-holo (red); (b): 2QWB-apo (black) and 2QWB-holo (red); (c): 3W09-int (blue).

turbation to its pK_as. In the following deglycosylation step, the neutralisation of the nearby nucleophile (glutamic acid) introduces a downshift in the pK_a of this catalytic residue and thus it becomes deprotonated to serve as a general base. Altogether, our studies suggest that the electrostatic perturbation upon substrate binding and glycosylation, coupled with the change in the local solvation environment, serve to precisely establish the required pK_as for its dual role during the hydrolysis. This study offers a complementary atomistic view of the pK_a cycling in GH33 and 34 and also provides a basis for future studies on glycoside hydrolyases and transferases, where there are substantial debates over their catalytic mechanisms.

5 Acknowledgement

We would like to acknowledge Dr Moritz Winger (Griffith University, Australia) and Dr Xiao Zhu (TACC, USA) for comments. This work was supported by UOW Small Grant Scheme. Haibo Yu is an Australian Research Council Future Fellow (FT110100034). This research was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government.

References

(1) G. S. Jacob, *Curr. Opin. Struct. Biol.*, 1995, **5**, 605–611.
(2) J. D. McCarter and S. G. Withers, *Curr. Opin. Struct. Biol.*, 1994, **4**, 885–892.
(3) G. Davies and B. Henrissat, *Structure*, 1995, **3**, 853–859.
(4) D. L. Zechel and S. G. Withers, *Acc. Chem. Res.*, 2000, **33**, 11–18.
(5) D. J. Vocadlo and G. J. Davies, *Curr. Opin. Chem. Biol.*, 2008, **12**, 539–555.
(6) V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490–5.
(7) D. E. Koshland, *Biol. Rev.*, 1953, **28**, 416–436.
(8) T. Angata and A. Varki, *Chem. Rev.*, 2002, **102**, 439–469.
(9) A. Varki, *FASEB J*, 1997, **11**, 248–255.
(10) A. G. Watts, I. Damager, M. L. Amaya, A. Buschiazzo, P. Alzari, A. C. Frasch and S. G. Withers, *J. Amer. Chem. Soc.*, 2003, **125**, 7532–7533.
(11) C. J. Vavricka, Y. Liu, H. Kiyota, N. Sriwilaijaroen, J. Qi, K. Tanaka, Y. Wu, Q. Li, Y. Li, J. Yan, Y. Suzuki and G. F. Gao, *Nat. Commun.*, 2013, **4**, 1491.
(12) J.-H. Kim, R. Resende, T. Wennekes, H.-M. Chen, N. Bance, S. Buchini, A. G. Watts, P. Pilling, V. A. Streltsov, M. Petric, R. Liggins, S. Barrett, J. L.

McKimm-Breschkin, M. Niikura and S. G. Withers, *Science*, 2013, **340**, 71–75.
(13) B. R. Miller and A. E. Roitberg, *Future Med. Chem.*, 2013, **5**, 1889–1900.
(14) M. von Itzstein, *Nat. Rev. Drug. Discov.*, 2007, 967–974.
(15) S. Schenkman, D. Eichinger, M. E. Pereira and V. Nussenzweig, *Ann. Rev. Microbiol.*, 1994, **48**, 499–523.
(16) P. M. Colman, *Prot. Sci.*, 1994, **3**, 1687–1696.
(17) G. Pierdominici-Sottile and A. E. Roitberg, *Biochemistry*, 2011, **50**, 836–842.
(18) G. Pierdominici-Sottile, N. A. Horenstein and A. E. Roitberg, *Biochemistry*, 2011, **50**, 10150–10158.
(19) L. P. McIntosh, G. Hand, P. E. Johnson, M. D. Joshi, M. Korner, L. A. Plesniak, L. Ziser, W. W. Wakarchuk and S. G. Withers, *Biochemistry*, 1996, **35**, 9958–9966.
(20) D. J. Vocadlo, G. J. Davies, R. Laine and S. G. Withers, *Nature*, 2001, **412**, 835–838.
(21) M. L. Ludwiczek, I. D'Angelo, G. N. Yalloway, J. A. Brockerman, M. Okon, J. E. Nielsen, N. C. J. Strynadka, S. G. Withers and L. P. McIntosh, *Biochemistry*, 2013, **52**, 3138–3156.
(22) S. M. Parsons and M. A. Raftery, *Biochemistry*, 1972, **11**, 1623–1629.
(23) S. M. Parsons and M. A. Raftery, *Biochemistry*, 1972, **11**, 1630–1633.
(24) A. Buschiazzo, M. F. Amaya, M. L. Cremona, A. C. Frasch and P. M. Alzari, *Mol. Cell.*, 2002, **10**, 757–768.
(25) J. N. Varghese, P. W. Smith, S. L. Sollis, T. J. Blick, A. Sahasrabudhe, J. L. McKimm-Breschkin and P. M. Colman, *Structure*, 1998, **6**, 735–746.
(26) M. F. Amaya, A. G. Watts, I. Damager, A. Wehenkel, T. Nguyen, A. Buschiazzo, G. Paris, A. C. Frasch, S. G. Withers and P. M. Alzari, *Structure*, **12**, 775–784.
(27) T. K. Harris and G. J. Turner, *IUBMB Life*, 2002, **53**, 85–98.
(28) A. Warshel, *Biochemistry*, 1981, **20**, 3167–3177.
(29) B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.
(30) A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
(31) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
(32) P. Schaefer, D. Riccardi and Q. Cui, *J. Chem. Phys.*, 2005, **123**, 014905.
(33) M. Nina, D. Beglov and B. Roux, *J. Phys. Chem. B*, 1997, **101**, 5239–5248.
(34) J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, **23**, 327–341.
(35) M. Gaus, A. Goez and M. Elstner, *J. Chem. Theory Comput.*, 2012, **9**, 338–354.
(36) A. Warshel and M. Levitt, *J. Mol. Biol.*, **103**, 227–249.
(37) H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.
(38) R. Lonsdale, K. E. Ranaghan and A. J. Mulholland, *Chem. Commun.*, 2010, **46**, 2354–2372.
(39) Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2000, **105**, 569–585.
(40) D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, P. König, G. Li, D. Xu, H. Guo, M. Elstner and Q. Cui, *J. Phys. Chem. B*, 2006, **110**, 6458–6469.
(41) M. Gaus, Q. Cui and M. Elstner, *WIREs: Comput. Mol.Sci.*, 2014, **4**, 49–61.
(42) M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260–7268.
(43) Y. Yang, H. Yu, D. York, Q. Cui and M. Elstner, *J. Phys. Chem. A*, 2007, **111**, 10861–10873.
(44) P. H. König, M. Hoffmann, T. Frauenheim and Q. Cui, *J. Phys. Chem. B*, 2005, **109**, 9082–9095.
(45) G. Li and Q. Cui, *J. Phys. Chem. B*, 2003, **107**, 14521–14528.
(46) S. C. L. Kamerlin, M. Haranczyk and A. Warshel, *J. Phys. Chem. B*, 2008, **113**, 1253–1272.
(47) C. R. Søndergaard and M. Olsson, *J. Chem. Theo. Comput.*, 2011, **7**, 2284–2295.
(48) N. Ghosh and Q. Cui, *J. Phys. Chem. B*, 2008, **112**, 8387–8397.
(49) P. Goyal, N. Ghosh, P. Phatak and M. Clemens, *J. Amer. Chem. Soc.*, 2011, **133**, 14981–14997.
(50) H. Yu, I. M. Ratheal, P. Artigas and B. Roux, *Nat. Struct. Mol. Biol.*, 2011, **18**, 1159–1163.
(51) P. Goyal, J. Lu, S. Yang, M. R. Gunner and Q. Cui, *Proc. Natl. Acad. Sci. USA*, 2013, **110**, 18886–18891.

(52) P. Scudder, J. P. Doom, M. Chuenkova, I. D. Manger and M. E. Pereira, *J. Biol. Chem.*, 1993, **268**, 9886–9891.
(53) L. V. Gubareva, M. J. Robinson, R. C. Bethell and R. G. Webster, *J. Virol.*, 1997, **71**, 3385–3390.
(54) C. Deshairs, N. Kessler, M. Aymard and G. A. Quash, *J. Gen. Virol.*, 1986, **67**, 409–418.
(55) J. L. McKimm-Breschkin, J. Williams, S. Barrett, K. Jachno, M. McDonald, P. G. Mohr, T. Saito and M. Tashiro, *J. Antimicrob. Chemother.*, 2013, **68**, 2210–2221.
(56) A. E. Mark and W. F. van Gunsteren, *J. Mol. Biol.*, 1994, **240**, 167–176.